

## RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.

Application Serial Number: 10/780,507  
Source: IFWO  
Date Processed by STIC: 11-26-04

# ***ENTERED***



IFWO

## RAW SEQUENCE LISTING

DATE: 11/26/2004

PATENT APPLICATION: US/10/780,507

TIME: 10:21:34

Input Set : A:\16336-13-2 seq listing.txt

Output Set: N:\CRF4\11262004\J780507.raw

```

3 <110> APPLICANT: MULLINS, James I.
4   RODRIGO, Allen G.
5   LEARN, Gerald H.
6   LI, Fusheng
7   NICKLE, David C.
8   JENSEN, Mark A.
10 <120> TITLE OF INVENTION: ANCESTRAL AND COT VIRAL SEQUENCES, PROTEINS AND IMMUNOGENIC
COMPOSITIONS
12 <130> FILE REFERENCE: 16336-001320US
14 <140> CURRENT APPLICATION NUMBER: 10/780,507
15 <141> CURRENT FILING DATE: 2004-02-17
17 <150> PRIOR APPLICATION NUMBER: US 10/204,204
18 <151> PRIOR FILING DATE: 2001-02-16
20 <150> PRIOR APPLICATION NUMBER: PCT/US01/05288
21 <151> PRIOR FILING DATE: 2001-02-16
23 <150> PRIOR APPLICATION NUMBER: US 60/183,659
24 <151> PRIOR FILING DATE: 2000-02-18
26 <150> PRIOR APPLICATION NUMBER: US 60/447,586
27 <151> PRIOR FILING DATE: 2003-02-14
29 <160> NUMBER OF SEQ ID NOS: 125
31 <170> SOFTWARE: PatentIn version 3.1
33 <210> SEQ ID NO: 1
34 <211> LENGTH: 2652
35 <212> TYPE: DNA
36 <213> ORGANISM: Artificial Sequence
38 <220> FEATURE:
39 <223> OTHER INFORMATION: Ancestral HIV-1 group M, subtype B, env sequence
41 <400> SEQUENCE: 1
42 atgcgcgtga agggcatccg caagaactac cagcacctgt ggcgctgggg caccatgctg      60
44 ctggggatgc tgatgatctg ctccgcggcc gagaaactgt gggtgaccgt gtactacggc      120
46 gtgcccgtgt ggaaggaggc caccaccacc ctgttctgcg ccagcgacgc caaggcttac      180
48 gacaccgagg tccacaacgt gtgggccacc cagcctgcg tgcccaccga cccaacccc      240
50 caggaggttg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa caacatggtg      300
52 gagcagatgc acgaggacat catcagcctg tgggaccaga gctgaagcc ctgcgtgaag      360
54 ttaaccccc tgtgcgtgac cctgaactgc accgacgacc tgcgcaccaa cgccaccaac      420
56 accaccaaca gcagcgccac caccaacacc accagcagcg gcggcgccac gatggagggc      480
58 gagaagggcg agatcaagaa ctgcagcttc aacgtgacca ccagcatccg cgacaagatg      540
60 cagaaggagt acgcccgtgt ctacaagctg gacgtggtgc ccatcgacaa cgacaacaac      600
62 aacaccaaca acaacaccag ctaccgcctc atcaactgca acaccagcgt gatcaccag      660
64 gctgccccca aggtgagctt cgagcccatc cccatccact actgcacccc cgccggcttc      720
66 gccatcctga agtgcaacga caagaagttc aacggcaccg gcccctgcac caacgtgagc      780
68 accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc      840
70 agcctggccg aggaggaggt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc      900
72 atcatcgtgc agctgaacga gacgctggag atcaactgca cgcgtcccaa caacaacacc      960

```

## RAW SEQUENCE LISTING

DATE: 11/26/2004

PATENT APPLICATION: US/10/780,507

TIME: 10:21:34

Input Set : A:\16336-13-2 seq listing.txt

Output Set : N:\CRF4\11262004\J780507.raw

```

74 cgcaagagca tccccatcgg cccctggccgc gccctgtacg ccaccggcaa gatcatcggc 1020
76 gacatccgcc agggccactg caacctgtcg cgagccaagt ggaacaacac cctgaagcag 1080
78 atcgtgacca agctgcgcga gcagttcggc aacaacaaga ccaccatcgt gttcaaccag 1140
80 agcagcgccg gcgaccccgga gatcgtgatg cacagcttca actgcggcgg cgaattcttc 1200
82 tactgcaaca gcacccagct gttcaacagc acctggcact tcaacggcac ctggggcaac 1260
84 aacaacaccg agcgcagcaa caacgcgcgc gacgacaacg acaccatcac cctgccctgc 1320
86 cgcataaagc agatcatcaa catgtggcag gaggtgggca aggccatgta cgcccccccc 1380
88 atcagcgccc agatccgctg cagcagcaac atcaccggcc tgctgctgac tcgagacggc 1440
90 ggcaacaacg agaacaccaa caacaccgac accgagatct tccgcccccg gggcggcgac 1500
92 atgcgcgaca actggcgagc cgagctgtac aagtacaagg tggatgaagat cgagccccctg 1560
94 ggcgtggccc ccaccaaggc caagcgccgc gtggtgcagc gcgagaagcg cgccgtgggc 1620
96 atgctggggc ccatgttctt gggcttcttg ggcgcgcgcg gcagcaccat gggcgccgcc 1680
98 agcatgaccc tgaccgtgca ggcccgcag ctgctgagcg gcacgtgca gcagcagaac 1740
100 aacctgctgc gcgccatcga ggcccagcag cacctgctgc agctgaccgt gtggggcatc 1800
102 aagcagctgc agggccgcgt gctggccgtg gagcggatcc tgaaggacca gcagctgctg 1860
104 ggcattctgg gctgcagcgg caagctgatc tgcaccaccg cggtgccctg gaacgccagc 1920
106 tggagcaaca agagcctgga caagatctgg aacaacatga cctggatgga gtgggagcgc 1980
108 gagatcgaca actacaccgg cctgatctac acctgatcg aggagagcca gaaccagcag 2040
110 gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 2100
112 gatataacca actggctgtg gtacatcaag atcttcatca tgatcgtggg cggcctggtg 2160
114 ggcctgcgca tcgtgttcgc cgtgctgagc atcgtgaacc gcgtgcgcca gggctacagc 2220
116 cccctgagct tccagaccgc cctgcccgcc cccgcggccc ccgaccgccc cgagggcac 2280
118 gaggaggagg gcggcgagcg cgaccgcgac cgacgcgggc gcctggtgaa cggcttcctg 2340
120 gccctgatct gggacgacct gcgcagcctg tgctgttca gctaccaccg cctgcgcgac 2400
122 ctgctgctga tcgtggcccg catcgtggag ctgctgggccc ggcgcggctg ggaggccctg 2460
124 aagtattggt ggaacctgct gcagtactgg agccaggagc tgaagaacag cgccgtgagc 2520
126 ctgctgaacg ccaccgccat cgccgtggcc gagggcaccg accgcgtgat cgaggtggtg 2580
128 cagcgcgcct gcgcgcgat cctgcacatc cccgcgcgca tccgccaggg cctggagcgc 2640
130 gccctgctgt ga 2652

```

133 &lt;210&gt; SEQ ID NO: 2

134 &lt;211&gt; LENGTH: 883

135 &lt;212&gt; TYPE: PRT

136 &lt;213&gt; ORGANISM: Artificial Sequence

138 &lt;220&gt; FEATURE:

139 &lt;223&gt; OTHER INFORMATION: Ancestral HIV-1 group M, subtype B, env sequence

141 &lt;400&gt; SEQUENCE: 2

```

143 Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp
144 1           5           10           15
147 Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Lys
148           20           25           30
151 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
152           35           40           45
155 Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
156           50           55           60
159 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
160 65           70           75           80
163 Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
164           85           90           95
167 Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp

```

## RAW SEQUENCE LISTING

DATE: 11/26/2004

PATENT APPLICATION: US/10/780,507

TIME: 10:21:34

Input Set : A:\16336-13-2 seq listing.txt

Output Set: N:\CRF4\11262004\J780507.raw

168																
171	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys	Val	Thr	Leu
172			115					120					125			
175	Asn	Cys	Thr	Asp	Asp	Leu	Arg	Thr	Asn	Ala	Thr	Asn	Thr	Thr	Asn	Ser
176		130					135					140				
179	Ser	Ala	Thr	Thr	Asn	Thr	Thr	Ser	Ser	Gly	Gly	Gly	Thr	Met	Glu	Gly
180	145					150					155					160
183	Glu	Lys	Gly	Glu	Ile	Lys	Asn	Cys	Ser	Phe	Asn	Val	Thr	Thr	Ser	Ile
184					165					170					175	
187	Arg	Asp	Lys	Met	Gln	Lys	Glu	Tyr	Ala	Leu	Phe	Tyr	Lys	Leu	Asp	Val
188				180					185					190		
191	Val	Pro	Ile	Asp	Asn	Asp	Asn	Asn	Asn	Thr	Asn	Asn	Asn	Thr	Ser	Tyr
192			195					200					205			
195	Arg	Leu	Ile	Asn	Cys	Asn	Thr	Ser	Val	Ile	Thr	Gln	Ala	Cys	Pro	Lys
196		210					215					220				
199	Val	Ser	Phe	Glu	Pro	Ile	Pro	Ile	His	Tyr	Cys	Thr	Pro	Ala	Gly	Phe
200	225					230					235				240	
203	Ala	Ile	Leu	Lys	Cys	Asn	Asp	Lys	Lys	Phe	Asn	Gly	Thr	Gly	Pro	Cys
204					245					250					255	
207	Thr	Asn	Val	Ser	Thr	Val	Gln	Cys	Thr	His	Gly	Ile	Arg	Pro	Val	Val
208				260					265					270		
211	Ser	Thr	Gln	Leu	Leu	Leu	Asn	Gly	Ser	Leu	Ala	Glu	Glu	Glu	Val	Val
212			275					280					285			
215	Ile	Arg	Ser	Glu	Asn	Phe	Thr	Asp	Asn	Ala	Lys	Thr	Ile	Ile	Val	Gln
216		290					295					300				
219	Leu	Asn	Glu	Ser	Val	Glu	Ile	Asn	Cys	Thr	Arg	Pro	Asn	Asn	Asn	Thr
220	305					310					315				320	
223	Arg	Lys	Ser	Ile	Pro	Ile	Gly	Pro	Gly	Arg	Ala	Leu	Tyr	Ala	Thr	Gly
224					325					330					335	
227	Lys	Ile	Ile	Gly	Asp	Ile	Arg	Gln	Ala	His	Cys	Asn	Leu	Ser	Arg	Ala
228				340					345					350		
231	Lys	Trp	Asn	Asn	Thr	Leu	Lys	Gln	Ile	Val	Thr	Lys	Leu	Arg	Glu	Gln
232			355					360					365			
235	Phe	Gly	Asn	Asn	Lys	Thr	Thr	Ile	Val	Phe	Asn	Gln	Ser	Ser	Gly	Gly
236		370					375					380				
239	Asp	Pro	Glu	Ile	Val	Met	His	Ser	Phe	Asn	Cys	Gly	Gly	Glu	Phe	Phe
240	385					390					395				400	
243	Tyr	Cys	Asn	Ser	Thr	Gln	Leu	Phe	Asn	Ser	Thr	Trp	His	Phe	Asn	Gly
244					405					410					415	
247	Thr	Trp	Gly	Asn	Asn	Asn	Thr	Glu	Arg	Ser	Asn	Asn	Ala	Ala	Asp	Asp
248				420												

## RAW SEQUENCE LISTING

DATE: 11/26/2004

PATENT APPLICATION: US/10/780,507

TIME: 10:21:34

Input Set : A:\16336-13-2 seq listing.txt

Output Set: N:\CRF4\11262004\J780507.raw

```

267 Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr
268           500           505           510
271 Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys
272           515           520           525
275 Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly Met Leu Gly Ala
276           530           535           540
279 Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala
280 545           550           555           560
283 Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val
284           565           570           575
287 Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu
288           580           585           590
291 Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu
292           595           600           605
295 Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly
296           610           615           620
299 Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser
300 625           630           635           640
303 Trp Ser Asn Lys Ser Leu Asp Lys Ile Trp Asn Asn Met Thr Trp Met
304           645           650           655
307 Glu Trp Glu Arg Glu Ile Asp Asn Tyr Thr Gly Leu Ile Tyr Thr Leu
308           660           665           670
311 Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu
312           675           680           685
315 Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Asn
316           690           695           700
319 Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Val
320 705           710           715           720
323 Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg
324           725           730           735
327 Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Arg Leu Pro Ala Pro Arg
328           740           745           750
331 Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp
332           755           760           765
335 Arg Asp Arg Ser Gly Arg Leu Val Asn Gly Phe Leu Ala Leu Ile Trp
336           770           775           780
339 Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp
340 785           790           795           800
343 Leu Leu Leu Ile Val Ala Arg Ile Val Glu Leu Leu Gly Arg Arg Gly
344           805           810           815
347 Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln
348           820           825           830
351 Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala Thr Ala Ile Ala
352           835           840           845
355 Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Val Gln Arg Ala Cys
356           850           855           860
359 Arg Ala Ile Leu His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg
360 865           870           875           880
363 Ala Leu Leu

```

## RAW SEQUENCE LISTING

DATE: 11/26/2004

PATENT APPLICATION: US/10/780,507

TIME: 10:21:34

Input Set : A:\16336-13-2 seq listing.txt

Output Set: N:\CRF4\11262004\J780507.raw

```

367 <210> SEQ ID NO: 3
368 <211> LENGTH: 2562
369 <212> TYPE: DNA
370 <213> ORGANISM: Artificial sequence
372 <220> FEATURE:
373 <223> OTHER INFORMATION: Ancestral HIV-1 group M, subtype C, env sequence
375 <400> SEQUENCE: 3
376 atgcggtgga tgggcatcct gcggaactgc cagcagtggg ggatctgggg catcctgggc      60
378 ttctggatgc tgatgatctg cagcgtgatg ggcaacctgt ggggtgaccgt gtactacggc      120
380 gtgcccgtgt ggaaggaggg caagaccacc ctgttctgcg ccagcgacgc caaggcctac      180
382 gagcgggagg tgcacaacgt gtgggccacc cagcctgcg tggccaccga ccccaacccc      240
384 caggagatgg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg      300
386 gaccagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgctgaag      360
388 ctgacccccc tgtgctgac cctgaactgc accaactgta ccaacaccaa caacaacaac      420
390 aacaccagca tgggcggcga gatcaagaac tgcagcttca acatcaccac cgagctgcgg      480
392 gacaagaagc agaaggtgta cgccctgttc taccgctggg acatcgtgcc cctgaacgag      540
394 aacagcaaca gcaacagcag cgagtaccgg ctgatcaact gcaacaccag cgccatcacc      600
396 caggcctgcc ccaaggtgag cttcgacccc atcccctcc actactgcgc ccccgccggc      660
398 tacgccatcc tgaagtgcaa caacaagacc ttcaacggca cgggccctg caacaacgtg      720
400 agcaccgtgc agtgaccca cggcatcaag ccgctggtga gcacccagct gctgctgaac      780
402 ggcagcctgg ccgaggagga gatcatcatc cggagcgaga acctgaccaa caacgccaag      840
404 accatcatcg tgcacctgaa cgagagcgtg gagatcgtgt gcacccggcc caacaacaac      900
406 acccggaaga gcatccggat cggccccggc cagaccttct acgccaccgg cgacatcatc      960
408 ggcgacatcc ggcaggccca ctgcaacatc agcgagaagg agtggaaaca gaccctgcag      1020
410 cgggtgggca agaagctgaa ggagcacttc cccaacaaga ccatcaagtt cgagcccagc      1080
412 agcggcgggc acctggagat caccaccac agcttcaact gccggggcga gttcttctac      1140
414 tgcaacacca gcaagctgtt caacagcacc tacaacagca ccaacaacgg caccaccagc      1200
416 aacagcacca tcacctgcc ctgcccgatc aagcagatca tcaacatgtg gcagggcgtg      1260
418 ggcggggcca tgtacgcccc ccccatcgcc ggcaacatca cctgcaagag caacatcacc      1320
420 ggcctgctgc tgaccggga cggcggaac accaacaaca ccaccgagac ctccggccc      1380
422 ggcggcgggc acatgcggga caactggcgg agcgagctgt acaagtacaa ggtggtggag      1440
424 atcaagcccc tgggcgtggc cccaccgag gccaaagcgg ggggtggtgga gcgggagaag      1500
426 cgggcctggt gcatcgcgcc cgtgttctct ggcttctggt gcgcggcgg cagcaccatg      1560
428 ggcgcgcca gcatcacctt gaccgtgcag gcccgccagc tgetgagcgg catcgtgcag      1620
430 cagcagagca acctgctgcg ggccatcgag gcccagcagc acatgctgca gctgaccgtg      1680
432 tggggcatca agcagctgca gaccggggtg ctggccatcg agcgggtacct gaaggaccag      1740
434 cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg      1800
436 aacagcagct ggagcaacaa gagccaggac gacatctggg acaacatgac ctggatgcag      1860
438 tgggaccggg agatcagcaa ctacaccgac accatctacc ggctgctgga ggacagccag      1920
440 aaccagcagg agaagaacga gaaggacctg ctggccctgg acagctggaa gaacctgtgg      1980
442 aactggttcg acatcaccaa ctggctgtgg tacatcaaga tcttcatcat gatcgtgggc      2040
444 ggcctgateg gectgcggat catcttcgcc gtgctgagca tegtgaaccg ggtgcggcag      2100
446 ggctacagcc cctgagctt ccagacctg accccaacc cccggggccc cgaccggctg      2160
448 ggcggcatcg aggaggagg cggcgagcag gaccgggacc ggagcatccg gctggtgagc      2220
450 ggttctctgg cctggcctg ggacgacctg cggagcctgt gcctgttcag ctaccaccgg      2280
452 ctgcgggaact tcatcctgat cgccgcccgg ggctgaacc tgetgggccc gagcagcctg      2340
454 cggggcctgc agcggggctg ggaggccctg aagtacctgg gcagcctggt gcagtactgg      2400
456 ggcctggagc tgaagaagag cgccatcagc ctgctggaca ccatcgccat cgccgtggcc      2460
458 gagggcaccg accggatcat cgagctggtg cagcggatct gccggggccat ccggaacatc      2520

```

RAW SEQUENCE LISTING ERROR SUMMARY

DATE: 11/26/2004

PATENT APPLICATION: US/10/780,507

TIME: 10:21:35

Input Set : A:\16336-13-2 seq listing.txt

Output Set: N:\CRF4\11262004\J780507.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 10

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/780,507

DATE: 11/26/2004

TIME: 10:21:35

Input Set : A:\16336-13-2 seq listing.txt

Output Set: N:\CRF4\11262004\J780507.raw